Two series of formalized interpretability principles for weak systems of arithmetic

Evan Goris, Joost J. Joosten

April 1, 2015

Abstract

The provability logic of a theory T captures the structural behavior of formalized provability in T as provable in T itself. Like provability, one can formalize the notion of relative interpretability giving rise to interpretability logics. Where provability logics are the same for all moderately sound theories of some minimal strength, interpretability logics do show variations.

The logic $\mathbf{IL}(\mathrm{All})$ is defined as the collection of modal principles that are provable in any moderately sound theory of some minimal strength. In this paper we raise the previously known lower bound of $\mathbf{IL}(\mathrm{All})$ by exhibiting two series of principles which are shown to be provable in any such theory. Moreover, we compute the collection of frame conditions for both series.

1 Introduction

Relative interpretations in the sense of Tarski, Mostowski and Robinson [10] are widely used in mathematics and in mathematical logic to interpret one theory into another. Roughly speaking, such an interpretation between two theories is a *translation* from the language of one theory to the language of the other so that the translation preserves logical structure and theoremhood.

We shall write $U \rhd V$ to denote that a theory U interprets a theory V. Once we know that $U \rhd V$, this provides us much information; for example the consistency of U implies the consistency of V and also, various definability results carry over from the one theory to the other. Famous examples of interpretations are abundant: the theory of the natural numbers into the theory of the integers, set theory plus the continuum hypothesis into ordinary set theory, non-Euclidean geometry into Euclidean geometry, etc.

Interpretability, being a syntactical notion, allows for formalization very much as one can formalize the notion of provability. As such, we can consider *interpretability logics* which actually extend the well-know provability logic **GL** of Gödel Löb. We shall see that the interpretability logic of a theory is the collection of all structural properties of interpretability that it can prove.

Where all modestly correct theories of some minimal strength –let us call them *reasonable theories* in this paper– have the same provability logic **GL**, the situation is different in the case of interpretability and different theories have different logics. It is an open question to determine the logic of interpretability principles being provable in any reasonable theory. This paper reports on substantial progress on this open question by increasing the previously known lower bound.

2 Preliminaries

Let U and V denote theories with languages $\mathcal{L}_U$ and $\mathcal{L}_V$ respectively. A relative interpretation j from V into a theory U –we will write $j : U \rhd V$– is a pair $\langle \delta(x), t\rangle$ where $\delta(x)$ is a formula of $\mathcal{L}_U$ that specifies the domain in which V will be interpreted and t is a translation, mapping symbols of $\mathcal{L}_V$ to formulas of $\mathcal{L}_U$ providing a definition in U of these symbols.

The translation t is extended to a translation j of formulas in the usual way by having j commute with the connectives and relativize the quantifiers to the domain specifier $\delta(x)$ as follows: $\big(\forall x\ \varphi(x)\big)^j := \forall x\ \big(\delta(x) \to \varphi^j(x)\big)$. We will not go too much into details but the main point is that interpretations are primarily syntactical notions –especially for finite languages– and as such allow for an arithmetization/formalization very much as formal proofs do.

2.1 Arithmetic

In order to formalize the notion of interpretability within some base theory T one needs to require some minimal strength conditions on T. In particular, we shall require that T can speak of numbers where to code syntax and without loss of generality we shall assume that T contains the language of arithmetic $\{+, \times, S, 0, 1 <, =\}$.

We will need that the main properties of the basic syntactical operations like substitution are provable within T. For reasonable coding protocols this implies that we need to require the totality of a function of growth-rate $\omega_1(x) := x \mapsto 2^{2|x|}$ where $|x|$ denotes the integral part of the binary logarithm of x.

Further, to perform basic arguments we need a minimal amount of induction and actually a surprisingly little amount of induction suffices. Buss's theory S^1_2 has just the needed amount of induction and proves the totality of ω_1 and this shall be our base theory (formulated in the standard language of arithmetic).

Alternatively, we could have taken as base theory $\mathrm{I}\Delta_0 + \Omega_1$ which consists of Robinson's arithmetic Q together with induction for bounded formulas with parameters and the axiom Ω_1 stating that the graph of ω_1 defines a total function. We refer the reader to [5] and [2] for further details.

A sharply bounded quantifier is one of the form $\forall\, x{<}|y|$ where $|y|$ denotes the integer value of the binary logarithm of y. The class Δ^b_0 contains exactly the formulas where each quantifier is sharply bounded. The class Σ^1_b arises by

allowing bounded existential quantifiers and sharply bounded universal quantifiers to occur over Δ_0^b formulas. By $\exists\Sigma_b^1$ we denote those formulas that arise by allowing a single unbounded existential quantifier over a Σ_b^1 formula. The complexity classes Π_n, Σ_n and Δ_n refer to the usual quantifier alternations hierarchies in the standard language of arithmetic.

In this paper we shall only be concerned with first order-theories in the language of arithmetic with a poly-time recognizable set of axioms extending S_2^1 and shall often refrain from repeating (some of) these conditions. We shall write $\Box_T\phi$ as the $\exists\Sigma_1^b$ formalization of ϕ being provable in the theory T and refrain from distinguishing formulas from their Gödel numbers or even the numerals thereof. It is well known that we can express provable Σ_1 completeness using formalized provability.

Lemma 2.1. *For any theory T extending S_2^1 we have that*

$$T \vdash \forall\alpha\ \Box_T\alpha \to \Box_T\Box_T\alpha.$$

We will use $U \rhd V$ to denote the formalization of "the theory V is interpretable in the theory U". If we abbreviate the existential quantifier over numbers that code a pair $\langle\delta(x), t\rangle$ defining an interpretation by $\exists^{int} j$ we can write

$$U \rhd V := \exists^{int} j\, \forall\psi\ (\Box_V\psi \to \Box_U\psi^j). \tag{1}$$

An interpretation $j : U \rhd V$ can be used as a uniform way to obtain a model of V inside any model of U. If U satisfies full induction, then we see that actually the defined model of V is an end extension of the model of U: we define $f(0) := 0$ and $f(x+1) := f(x) +^j 1^j$ and by induction see that $\forall x \exists y\ f(x) = y$. As such, we see that any Σ_1 consequence of U must necessarily also hold in V. Since $\Box_T\varphi$ is a Σ_1 formula, the insight on end extensions is reflected in what is called *Montagna's principle*

$$(U \rhd V) \to \big((U \cup \{\Box_T\varphi\}) \rhd (V \cup \{\Box_T\varphi\})\big). \tag{2}$$

In case U does not have full induction, we can still define the graph $F(x, y)$ of the function f from above, but we can no longer prove that the function is total. However, we can prove that $\exists y\ F(x, y)$ is *progressive*, that is, we can prove

$$\exists y\, F(0, y)\ \wedge\ \forall x\ \big(\exists y\, F(x, y) \to \exists y,\ F(x+1, y)\big).$$

In particular, the formula $\exists y\, F(x, y)$ defines an initial segment within U. A common trick in weak arithmetics is to use this initial segment as our natural numbers instead of applying induction (which is not necessarily available). By Solovay's techniques on shortening initial segments we may assume that they obey certain closure properties giving rise to the what is called a *definable cut*.

A formula J is called a T-cut whenever T proves all of

1. $J(0) \wedge \forall x\, (J(x) \to J(x+1))$;
2. $\forall x\ \big(J(x) \wedge J(y) \to J(x+y) \wedge J(xy) \wedge J(\omega_1(x))\big)$;

3. $J(x) \wedge y < x \to J(y)$.

Let $\mathsf{Cut}(J)$ denote the conjunction of these three requirements. Sometimes we want to quantify over cuts within T so that these cuts can then of course be non-standard. We shall use $\forall^{\mathsf{Cut}} J\ \psi$ and $\exists^{\mathsf{Cut}} J\ \psi$ to denote $\forall J\ (\Box_T \mathsf{Cut}(\dot{J}) \to \psi)$ and $\exists J\ (\Box_T \mathsf{Cut}(\dot{J}) \wedge \psi)$ respectively. Here the dot notation in $\Box_T \mathsf{Cut}(\dot{J})$ is the standard way to abbreviate the formula with one free variable J stating that the formula $\mathsf{Cut}(J)$ is provable in T. We will freely use the dot notation throughout the remainder of this paper. Sometimes we shall write $x{\in}J$ instead of $J(x)$.

For J a cut, let ψ^J denote the formula where all unrestricted quantifiers are now required to range over values in J. That is, $(\forall x\ \phi)^J := \forall x\ (J(x) \to \phi^J)$, $(\exists x\ \phi)^J := \exists x\ (J(x) \wedge \phi^J)$. Moreover, that is the only thing that is done by this translation so that for example $(\phi \wedge \psi)^J := \phi^J \wedge \psi^J$ etc. Instead of writing $(\Box_T \phi)^J$ we shall simply write $\Box_T^J \phi$. We note that if $\psi(J) \in \exists\Sigma_1^b$, then $\exists^{\mathsf{Cut}} J\ \psi(J)$ is again provably equivalent to an $\exists\Sigma_1^b$ formula.

Let us get back to the role of induction in Montagna's principle. If $j : U \rhd V$ and U does not prove full induction, then j will not define an end extension of any model of U. However, it is easy to see that j does define, using the progressive formula $\exists y\, F(x,y)$, a definable cut in U on which f is an isomorphism. This is reflected in a weakening of Montagna's principle also referred to as *Pudlák's principle*.

Lemma 2.2. *Let T be a theory containing S_2^1 and let U and V be theories.*

$$T \vdash U \rhd V \to \exists^{\mathsf{Cut}} J\, \forall\, \psi{\in}\Delta_0\, \Big(U \cup \{(\exists x\, \psi)^{\dot{J}}\} \rhd V \cup \{\exists x\, \psi\}\Big). \tag{3}$$

2.2 The interpretability logic of a theory

Interpretability logics are designed to capture structural behavior of formalized interpretability just as provability logic captures the structural behavior of formalized provability. To this end we consider a propositional modal language with a unary modal operator $\Box$ to model formalized provability and a binary modal operator $\rhd$ to model formalized interpretability of sentential extensions of some base theory. Let us make this more precise.

Let us fix an arithmetical theory T; By $*$ we will denote a *realization*, that is, any mapping from the set of propositional variables to sentences of T. The map $*$ is extended to the set of all modal formulas of interpretability logics as follows

$$\begin{array}{lcll}
(\neg A)^* & := & \neg A^* & \\
(A \wedge B)^* & := & A^* \wedge B^* & \text{and likewise for the other connectives} \\
(\Box A)^* & := & \Box_T A^* & \\
(A \rhd B)^* & := & (T + A^*) \rhd (T + B^*). &
\end{array}$$

We can now define the interpretability logic of a theory as those modal principles which are provable under any realization. With some liberal notation this is captured in the following.

Definition 2.3. *Let T be a theory containing S^1_2. We define the* interpretability logic of T *as*

$$\mathbf{IL}(\mathrm{T}) \ := \ \{A \mid \forall * \ T \vdash A^*\}.$$

Further, we define the interpretability logic of all arithmetical theories extending S^1_2 by

$$\mathbf{IL}(\mathrm{All}) \ := \ \{A \mid \forall T \, \forall * \ T \vdash A^*\}.$$

As a direct corollary to (2) –Montagna's principle– we can conclude that

$$(A \rhd B) \to \big((A \wedge \Box C) \rhd (B \wedge \Box C)\big) \in \mathbf{IL}(\mathrm{T})$$

whenever T proves full induction. However, there is no direct reflection of Pudlák's principle on the level of interpretability logics since Pudlák's principle would translate to

$$(A \rhd B) \to \big((A \wedge \Box^J C) \rhd (B \wedge \Box C)\big) \in \mathbf{IL}(\mathrm{T})$$

for the particular cut J corresponding to $j : A \rhd B$ and this cannot be expressed in our modal language. In a sense, $\Box^J C$ corresponds to finding a small witness of the provability of C. As we shall see, there are various occasions where we can conclude that such small witnesses exist. The two main ingredients in obtaining such small witnesses are expressed by the following lemmas.

Lemma 2.4 (Outside big, inside small)**.** *For T, U any theories extending S^1_2, we have that*

$$T \vdash \forall^{\mathsf{Cut}} J \, \forall x \ \Box_U (\dot{x} \in \dot{J}).$$

Proof. Given J and given x, not necessarily in J, we can construct a proof-object to the extent that $x \in J$ in the obvious way. First conclude $J(0)$ which holds since J is a cut. Next, conclude that $J(1)$ from the progressiveness of J and $J(0)$ and so all the way to $J(x)$. This proof object is not much bigger than x itself. However it requires the totality of exponentiation. If this is not provable in T, the proof can be generalized by switching do dyadic numerals and we refer to e.g. [2, 7] for details. □

Lemma 2.5 (Formalized Henkin construction)**.** *For theories T, U and V all extending S^1_2 we have*

$$T \vdash \forall^{\mathsf{Cut}} J \ (U \cup \{\mathsf{Con}^J(V)\} \rhd V).$$

Proof. (Sketch) The theory T can verify that the usual Henkin construction can be formalized in U without many problems where J plays the role of the natural numbers. Instead of applying induction to obtain a maximal consistent set $\mathcal{M}_V$ as a consistent branch of infinite length in Lindenbaum's lemma, we can now only conclude that the length of the branch is within some cut I which is a shortening of J thereby yielding a set $\mathcal{M}^I_V$ which is contradiction-free on I.

The set $\mathcal{M}^I_V$ can be used to obtain a term model and we define an interpretation $j : (U \cup \{\mathsf{Con}^J(V)\} \rhd V)$ from the term model as usual so that provably

$\phi^j \leftrightarrow \big(\phi \in \mathcal{M}_V^I\big)$. Note that since the interpretation of identity can be any equivalence relation, there is no need to move to equivalence classes in the construction of our term model. By construction we have $\Box_U \forall \phi \, \big(\mathsf{Con}^J(V) \wedge \Box_V^I \phi \rightarrow \phi^j\big)$. By the outside big, inside small principle and the formalized deduction theorem we now conclude that

$$\forall \phi \; (\Box_V \varphi \rightarrow \Box_{U \cup \{\mathsf{Con}^J(V)\}} \varphi)$$

which, by (1) is nothing but $(U \cup \{\mathsf{Con}^J(V)\} \rhd V)$. We refer to [13] where one can see that the necessary induction for this argument is available in S^1_2. $\Box$

Using these lemmas we can infer in various occasions the existence of small witnesses to provability.

Lemma 2.6. *For any theory T we have $T \vdash \neg(A \rhd \neg C) \rightarrow \forall^{\mathsf{Cut}} K \Diamond (A \wedge \Box^{\dot{K}} C)$.*

Proof. Reason in arbitrary T by contraposition and apply the Henkin construction on a cut. $\Box$

As a corollary to this lemma, we see that $(A \rhd B) \rightarrow \big(\neg(A \rhd \neg C) \rhd (B \wedge \Box C)\big) \in$ **IL**(T) for any T extending S^1_2. It is an open problem to classify the modal principles that hold in any theory extending S^1_2. This paper raises the previously known lower bound.

We formulate some other direct corollaries of the outside-big inside-small principle in the following useful lemma.

Lemma 2.7. *Let T be any theory containing S^1_2. We have that*

1. $T \vdash \forall A \; (\Box \dot{A} \rightarrow \forall^{\mathsf{Cut}} K \; \Box \Box^{\dot{K}} \dot{A})$;
2. $T \vdash \sigma \rightarrow \forall^{\mathsf{Cut}} K \; \Box \sigma^{\dot{K}}$ *for any formula σ in $\exists \Sigma^b_1$;*
3. $T \vdash \forall \, C {\in} \exists \Sigma^b_1 \, \forall^{\mathsf{Cut}} J \; (\exists x \, C \rightarrow \Box \, \exists \, x {\in} \dot{J} \, C)$.

One ingredient in proving interpretability principles arithmetically sound, is to find small witnesses. Another ingredient tells us how we can *keep* these witnesses small. A simple generalization of Pudlák's lemma which was first proved in [6] and tells us how to do so.

Lemma 2.8. *If $j : \alpha \rhd \beta$ then, for every cut I there exists a definable cut J such that for every γ we have that*

$$T \vdash \forall^{\mathsf{Cut}} I \, \exists^{\mathsf{Cut}} J \, \exists^{\mathsf{int}} j \, \Big(j : (\alpha \wedge \Box^J \gamma) \rhd (\beta \wedge \Box^I \gamma)\Big).$$

2.3 Modal interpretability logics

When working in interpretability logic, we shall adopt a reading convention that will allow us to omit many brackets. Thus, we say that the strongest binding 'connectives' are $\neg$, $\Box$ and $\Diamond$ which all bind equally strong. Next come $\wedge$ and $\vee$, followed by $\rhd$ and the weakest connective is $\rightarrow$. Thus, for example, $A \rhd B \rightarrow A \wedge \Box C \rhd B \wedge \Box C$ will be short for $(A \rhd B) \rightarrow \big((A \wedge \Box C) \rhd (B \wedge \Box C)\big)$.

If we do not disambiguate a formula of nested conditionals ($\rightarrow$ or $\rhd$), then this should be read as a conjunction. For example, $A \rhd B \rhd C$ should be read as $(A \rhd B) \wedge (B \rhd C)$ and likewise for implications.

We first define the core logic **IL** which shall be present in any other interpretability logic. As before, we work in a propositional signature where apart from the classical connectives we have a unary modal operator $\Box$ and a binary modal operator $\rhd$.

Definition 2.9 (IL). *The logic* **IL** *contains apart from all propositional logical tautologies, all instantiations of the following axiom schemes.*

L1 $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$

L2 $\Box A \rightarrow \Box\Box A$

L3 $\Box(\Box A \rightarrow A) \rightarrow \Box A$

J1 $\Box(A \rightarrow B) \rightarrow A \rhd B$

J2 $(A \rhd B) \wedge (B \rhd C) \rightarrow A \rhd C$

J3 $(A \rhd C) \wedge (B \rhd C) \rightarrow A \vee B \rhd C$

J4 $A \rhd B \rightarrow (\Diamond A \rightarrow \Diamond B)$

J5 $\Diamond A \rhd A$

The rules of the logic are Modus Ponens (from $A \rightarrow B$ and A, conclude B) and Necessitation (from A conclude $\Box A$).

It is not hard to see that $\mathbf{IL} \subseteq \mathbf{IL}(\mathrm{All})$. By $\mathbf{IL}\mathsf{M}$ we denote the logic that arises by adding Montagna's axiom scheme

$$\mathsf{M} : \quad A \rhd B \rightarrow A \wedge \Box C \rhd B \wedge \Box C$$

to **IL**. It follows from our earlier observations that $\mathbf{IL}\mathsf{M} \subseteq \mathbf{IL}(\mathrm{T})$ and the other inclusion can be proven too.

Theorem 2.10 (Berarducci [1], Shavrukov [9])**.** *If T proves full induction, then* $\mathbf{IL}(\mathrm{T}) = \mathbf{IL}\mathsf{M}$.

The logic $\mathbf{IL}\mathsf{P}$ arises by adding the axiom scheme

$$\mathsf{P}\ :\qquad A \rhd B \to \Box(A \rhd B)$$

to the basic logic $\mathbf{IL}$. If T is finitely axiomatizable it is easy to see that (1) is provably equivalent to a Σ_1 formula so that by provable Σ_1 completeness we see that $\mathbf{IL}\mathsf{P} \subseteq \mathbf{IL}(\mathrm{T})$ for any finitely axiomatized theory T that proves that exponentiation is a total function. If T can moreover prove the totality of superexponentiation `supexp` then the inclusion can be reversed too. Here, $\mathtt{supexp}(x)$ is defined as $x \mapsto 2^x_x$ with $2^n_0 := n$ and $2^n_{m+1} := 2^{(2^n_m)}$.

Theorem 2.11 (Visser [12])**.** *If T is finitely axiomatizable and proves the totality of* `supexp`*, then* $\mathbf{IL}(\mathrm{T}) = \mathbf{IL}\mathsf{P}$.

It follows that $\mathbf{IL} \subseteq \mathbf{IL}(\mathrm{All}) \subseteq (\mathbf{IL}\mathsf{P} \cap \mathbf{IL}\mathsf{M})$. In this paper we shall focus on these bounds.

2.4 Relational semantics

We can equip interpretability logics with a natural relational semantics often referred to as Veltman semantics.

Definition 2.12. *A* Veltman frame *is a triple $\langle W, R, \{S_x\}_{x\in W}\rangle$ where W is a non-empty set of* possible worlds, *R a binary relation on W so that R^{-1} is transitive and well-founded. The $\{S_x\}_{x\in W}$ is a collection of binary relations on $x\uparrow$ (where $x\uparrow := \{y \mid xRy\}$). The requirements are that the S_x are reflexive and transitive and the restriction of R to $x\uparrow$ is contained in S_x, that is $R \cap (x\uparrow) \subseteq S_x$.*

A Veltman model *consists of a Veltman frame together with a valuation $V : \mathtt{Prop} \to \mathcal{P}(W)$ that assigns to each propositional variable $p \in \mathtt{Prop}$ a set of worlds $V(p)$ in W where p is stipulated to be true. This valuation defines a forcing relation $\Vdash \subseteq W \times \mathsf{Form}$ telling us which formulas are true at which particular world:*

$$\begin{array}{lll}
x \Vdash \bot & & \text{for no } x \in W;\\
x \Vdash A \to B & :\Leftrightarrow & x \nVdash A \text{ or } x \Vdash B;\\
x \Vdash \Box A & :\Leftrightarrow & \forall y\ (xRy \to y \Vdash A);\\
x \Vdash A \rhd B & :\Leftrightarrow & \forall y\ \Big(xRy \wedge y \Vdash A \to \exists z\ (yS_xz \wedge z \Vdash B\Big).
\end{array}$$

For a Veltman model $\mathcal{M} = \langle W, R, \{S_x\}_{x\in W}, V\rangle$, we shall write $\mathcal{M} \models A$ as short for $\forall\, x{\in}W\ \mathcal{M}, x \Vdash A$.

The logic $\mathbf{IL}$ is sound and complete with respect to all Veltman models ([3]). Often one is interested in considering all models that can be defined over a frame. Thus, given a frame $\mathcal{F}$ and a valuation V on $\mathcal{F}$ we shall denote the corresponding model by $\langle \mathcal{F}, V\rangle$. A *frame condition* for a modal formula P is a formula F (first or higher-order) in the language $\{R, \{S_x\}_{x\in W}\}$ so that $\mathcal{F} \models F$ (as a relational structure) if and only if $\forall^{\mathsf{valuation}} V\ \langle \mathcal{F}, V\rangle \models \mathsf{P}$.

It is easy to establish that the frame condition for P is $xRyRzS_xu \to zS_yu$ where $xRyRzS_xu$ is short for $xRy \wedge yRz \wedge zS_xu$. Likewise, it is elementary to see that the frame condition for M is given by $yS_xzRu \to yRu$. In this paper we shall compute the frame conditions for two new series of principles in **IL**(All).

Often we shall denote a valuation V directly by the induced forcing relation $\Vdash$. Given a Veltman model $\langle \mathcal{F}, \Vdash \rangle$ we define a *C-assuring successor* –denoted by $R^C_{\Vdash}$– as follows

$$xR^C_{\Vdash}y \quad := \quad \big(xRy \wedge y \Vdash C \;\wedge\; \forall z\ (yS_xz \to z \Vdash C)\big).$$

3 A slim hierarchy of principles

In this section we present a hierarchy of interpretability principles in **IL**(All) of growing strength. For a well-behaved sub-hierarchy we shall compute the frame conditions and prove arithmetical soundness. There is no particular 'slimness' inherent to the hierarchy presented here. The main reason for our name is that we tend to depict the frame conditions (see Figure 1) in a slim way as opposed to the depicted frame conditions for the series of principles that we refer to as a broad series of principles (see Figure 2).

3.1 A slim hierarchy

Inductively, we define a series of principles as follows.

$$\begin{array}{lll}
\mathsf{R_0} & := & A_0 \rhd B_0 \to \neg(A_0 \rhd \neg C_0) \rhd B_0 \wedge \Box C_0 \\[1em]
\mathsf{R_{2n+1}} & := & R_{2n}[\neg(A_n \rhd \neg C_n)/\neg(A_n \rhd \neg C_n) \wedge (E_{n+1} \rhd \Diamond A_{n+1}); \\
 & & B_n \wedge \Box C_n / B_n \wedge \Box C_n \wedge (E_{n+1} \rhd A_{n+1})] \\[1em]
\mathsf{R_{2n+2}} & := & R_{2n+1}[B_n / B_n \wedge (A_{n+1} \rhd B_{n+1}); \\
 & & \Diamond A_{n+1} / \neg(A_{n+1} \rhd \neg C_{n+1}); \\
 & & (E_{n+1} \rhd A_{n+1}) / (E_{n+1} \rhd A_{n+1}) \wedge (E_{n+1} \rhd B_{n+1} \wedge \Box C_{n+1})]
\end{array}$$

As to illustrate how these substitutions work we shall calculate the first five principles.

$$\begin{array}{lll}
\mathsf{R_0} & := & A_0 \rhd B_0 \to \neg(A_0 \rhd \neg C_0) \rhd B_0 \wedge \Box C_0 \\
\mathsf{R_1} & := & A_0 \rhd B_0 \to \neg(A_0 \rhd \neg C_0) \wedge (E_1 \rhd \Diamond A_1) \rhd B_0 \wedge \Box C_0 \wedge (E_1 \rhd A_1) \\
\mathsf{R_2} & := & A_0 \rhd B_0 \wedge (A_1 \rhd B_1) \to \neg(A_0 \rhd \neg C_0) \wedge (E_1 \rhd \neg(A_1 \rhd \neg C_1)) \rhd \\
 & & \qquad B_0 \wedge (A_1 \rhd B_1) \wedge \Box C_0 \wedge (E_1 \rhd A_1) \wedge (E_1 \rhd B_1 \wedge \Box C_1) \\
\mathsf{R_3} & := & A_0 \rhd B_0 \wedge (A_1 \rhd B_1) \to \\
 & & \neg(A_0 \rhd \neg C_0) \wedge (E_1 \rhd \neg(A_1 \rhd \neg C_1) \wedge (E_2 \rhd \Diamond A_2)) \rhd \\
 & & \quad B_0 \wedge (A_1 \rhd B_1) \wedge \Box C_0 \wedge (E_1 \rhd A_1) \wedge (E_1 \rhd B_1 \wedge \Box C_1 \wedge (E_2 \rhd A_2)) \\
\mathsf{R_4} & := & A_0 \rhd B_0 \wedge (A_1 \rhd B_1 \wedge (A_2 \rhd B_2)) \to \\
 & & \neg(A_0 \rhd \neg C_0) \wedge (E_1 \rhd \neg(A_1 \rhd \neg C_1) \wedge (E_2 \rhd \neg(A_2 \rhd \neg C_2))) \rhd \\
 & & \quad B_0 \wedge (A_1 \rhd B_1 \wedge (A_2 \rhd B_2)) \wedge \Box C_0 \wedge (E_1 \rhd A_1) \wedge \\
 & & \quad \big(E_1 \rhd B_1 \wedge (A_2 \rhd B_2) \wedge \Box C_1 \wedge (E_2 \rhd A_2) \wedge (E_2 \rhd B_2 \wedge \Box C_2)\big)
\end{array}$$

It is easy to see that the hierarchy defines a series of principles of increasing strength as expressed by the following lemma.

Lemma 3.1. *For each natural number n we have that* $\mathbf{IL}\mathsf{R_{n+1}} \vdash \mathsf{R_n}$.

Proof. By an easy case distinction. We see that $\vdash_{\mathsf{IL}} \mathsf{R}_{2n+1} \to \mathsf{R}_{2n}$ by choosing $E_{n+1} := \Diamond\top$ and $A_{n+1} := \top$. To see that $\vdash_{\mathsf{IL}} \mathsf{R}_{2n+2} \to \mathsf{R}_{2n+1}$ we choose $C_{n+1} := \top$ and $B_{n+1} := A_{n+1}$. $\Box$

Thus, to understand the hierarchy well, it suffices to study a well-behaved co-final subsequence of it. To this end we define the following hierarchy.

For any $n \geq 0$ we define schemata X_n, Y_n and Z_n as follows.

$$\begin{aligned}
\mathsf{X}_0 &= A_0 \rhd B_0;\\
\mathsf{Y}_0 &= \neg(A_0 \rhd C_0);\\
\mathsf{Z}_0 &= B_0 \wedge \Box C_0;
\end{aligned}$$

$$\begin{aligned}
\mathsf{X}_{n+1} &= A_{n+1} \rhd B_{n+1} \wedge (\mathsf{X}_n);\\
\mathsf{Y}_{n+1} &= \neg(A_{n+1} \rhd \neg C_{n+1}) \wedge (E_{n+1} \rhd \mathsf{Y}_n);\\
\mathsf{Z}_{n+1} &= B_{n+1} \wedge (\mathsf{X}_n) \wedge \Box C_{n+1} \wedge (E_{n+1} \rhd A_n) \wedge (E_{n+1} \rhd \mathsf{Z}_n).
\end{aligned}$$

For any $n \geq 0$ define

$$\widetilde{\mathsf{R}}_n = \mathsf{X}_n \to \mathsf{Y}_n \rhd \mathsf{Z}_n.$$

To see how this proceeds, let us evaluate the first couple of instances:

$$\begin{array}{lll}
\widetilde{\mathsf{R}}_0 & := & A_0 \rhd B_0 \to \neg(A_0 \rhd \neg C_0) \rhd B_0 \wedge \Box C_0;\\
\widetilde{\mathsf{R}}_1 & := & A_1 \rhd B_1 \wedge (A_0 \rhd B_0) \to\\
& & \quad \neg(A_1 \rhd \neg C_1) \wedge (E_1 \rhd \neg(A_0 \rhd \neg C_0)) \rhd\\
& & \qquad\qquad B_1 \wedge (A_0 \rhd B_0) \wedge \Box C_1 \wedge (E_1 \rhd A_0) \wedge (E_1 \rhd B_0 \wedge \Box C_0);\\
\widetilde{\mathsf{R}}_2 & := & A_2 \rhd B_2 \wedge (A_1 \rhd B_1 \wedge (A_0 \rhd B_0)) \to\\
& & \quad \neg(A_2 \rhd \neg C_2) \wedge \big(E_2 \rhd \neg(A_1 \rhd \neg C_1) \wedge (E_1 \rhd \neg(A_0 \rhd \neg C_0))\big) \rhd\\
& & \qquad\quad B_2 \wedge (A_1 \rhd B_1 \wedge (A_0 \rhd B_0)) \wedge \Box C_2 \wedge (E_2 \rhd A_1) \wedge\\
& & \qquad\quad \big(E_2 \rhd B_1 \wedge (A_0 \rhd B_0) \wedge \Box C_1 \wedge (E_1 \rhd A_0) \wedge (E_1 \rhd B_0 \wedge \Box C_0)\big);
\end{array}$$

It is clear that the $\widetilde{\mathsf{R}}_k$ hierarchy is directly related to the R_k hierarchy:

Lemma 3.2. *For each natural number k we have* $\mathsf{R}_{2k} := \widetilde{\mathsf{R}}_k[\mathbb{X}_i/\mathbb{X}_{k-i};\ E_i/E_{k+1-i}]$, *where* $\mathbb{X} \in \{A, B, C\}$.

Proof. By visual inspection we see that it holds for $k = 0, 1$. It is proven in full generality by an easy induction. To prove the lemma, it is best to consider the place-holders like A_i etc. as propositional variables since otherwise in principle, for example, A_i could contain E_i as a subformula. $\Box$

For the remainder of this section, we shall focus on the $\widetilde{\mathsf{R}}_k$ hierarchy and begin by computing a collection of frame condititions.

3.2 Frame conditions

For any $n \geq 0$ we define a ternary relation $\mathcal{G}_n(x, y, z)$ on Veltman-frames as follows.

$$\mathcal{G}_0(x, y, z) = \forall u\,(zRu \Rightarrow yS_x u),$$
$$\mathcal{G}_{n+1}(x, y, z) = \forall u\,\big(zRu \Rightarrow yS_x u \wedge \forall v\,(uS_x v \Rightarrow \mathcal{G}_n(z, u, v)\big).$$

For every $n \geq 0$ we define the first-order frame condition $\mathcal{F}_n$ as follows.

$$\mathcal{F}_n = \forall\, w, x, y, z\ (wRxRyS_w z \Rightarrow \mathcal{G}_n(x, y, z)).$$

The main result of this subsection is that $\mathcal{F}_{2n}$ is the frame correspondence of $\widetilde{\mathsf{R}}_n$. For $n = 0$ this has been established in [4]. It is easy to see that $\mathcal{G}_{n+1}(x, y, z)$ implies $\mathcal{G}_n(x, y, z)$ so that $\mathcal{F}_{n+1}$ also implies $\mathcal{F}_n$. The frame conditions $\mathcal{F}_k$ are depicted in Figure 1 for the first three values of k.

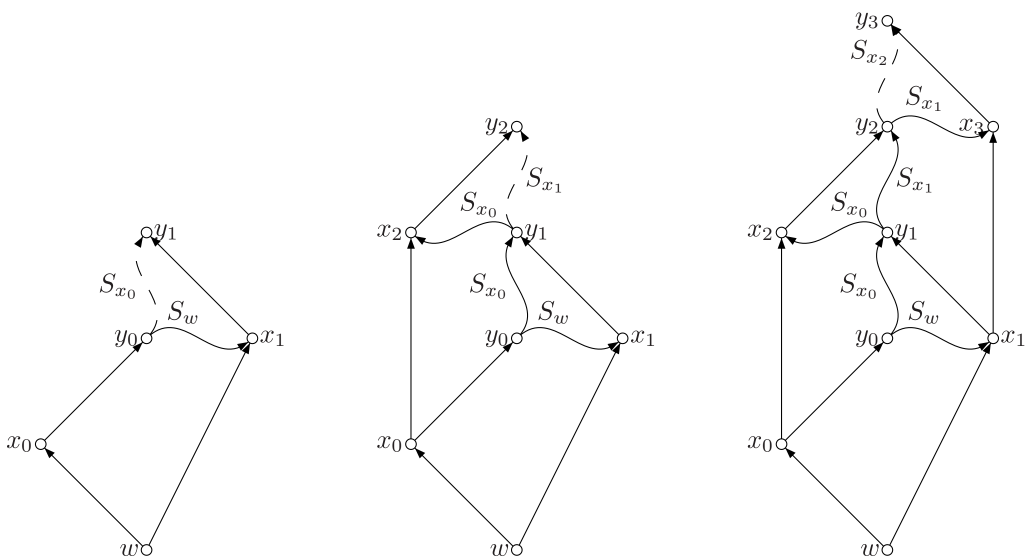

Figure 1: From left to right we have depicted $\mathcal{F}_0$ to $\mathcal{F}_2$. Since $\mathcal{F}_{k+1}$ implies $\mathcal{F}_k$ we have only depicted the content of $\mathcal{F}_{k+1}$ which is new w.r.t. $\mathcal{F}_k$. As such we should read the pictures as: “if all un-dashed relations are as in the picture, then also the dashed relation should be present”.

In what follows we let $F = \langle W, R, S \rangle$ be an arbitrary Veltman-frame. With a forcing relation $\Vdash$ we will always mean a forcing relation on F. For our convenience we define

$$A_{-1} \equiv \mathsf{X}_{-1} \equiv \mathsf{Z}_{-1} \equiv \top.$$

Before we can prove a frame correspondence we first need a technical lemma.

Lemma 3.3. *For all $k \geq 0$ and all $x, y, z \in W$. If $\mathcal{G}_{2k}(x, y, z)$ then for any forcing relation $\Vdash$ for which*

$$x \Vdash \mathsf{Y}_k \quad \textit{and} \quad xR_{\Vdash}^{C_k}y \quad \textit{and} \quad z \Vdash \mathsf{X}_{k-1},$$

we also have

$$z \Vdash \Box C_k \wedge (E_k \rhd A_{k-1}) \wedge (E_k \rhd \mathsf{Z}_{k-1}).$$

Proof. We shall write $xR^{C_k}y$ as short for $xR_{\Vdash}^{C_k}y$ and prove the claim by induction on k. With the convention that $A_{-1} \equiv \mathsf{Z}_{-1} \equiv \top$ the lemma is trivial for $k = 0$. So assume $k > 0$. Let $\Vdash$ be a forcing relation and take x, y and z such that

$$x \Vdash \neg(A_k \rhd \neg C_k) \wedge (E_k \rhd \mathsf{Y}_{k-1}), \tag{4}$$

$$xR^{C_k}y, \tag{5}$$

$$z \Vdash \mathsf{X}_{k-1}, \tag{6}$$

$$\mathcal{G}_{2k}(x, y, z). \tag{7}$$

Take an arbitrary $u \in W$ with zRu. By (7) we have yS_xu and thus by (5) we have $u \Vdash C_k$. This shows $z \Vdash \Box C_k$.

To show that also the other two conjuncts hold at z assume that $u \Vdash E_k$. By (4) we find some v with uS_xv and

$$v \Vdash \mathsf{Y}_{k-1}. \tag{8}$$

In order to show $z \Vdash E_k \rhd A_{k-1}$ we have to find some a with $uS_za \Vdash A_{k-1}$. Remark that Y_{k-1} implies $\Diamond A_{k-1}$ thus there exists some a with $vRa \Vdash A_{k-1}$. By (7) we have $\mathcal{G}_{2k-1}(z, u, v)$ and thus uS_za.

In order to show that also $z \Vdash E_k \rhd \mathsf{Z}_{k-1}$ we have to find some b with $uS_zb \Vdash \mathsf{Z}_{k-1}$. We just used that Y_{k-1} implies $\Diamond A_{k-1}$, but remark that Y_{k-1} implies the stronger statement that $\neg(A_{k-1} \rhd \neg C_{k-1})$. Thus there exists some a with $a \Vdash A_{k-1}$ and

$$vR^{C_{k-1}}a. \tag{9}$$

As above, by (7) we have $\mathcal{G}_{2k-1}(z, u, v)$ and thus uS_za and zRa. By (6) there exists a b with aS_zb and

$$b \Vdash B_{k-1} \wedge (\mathsf{X}_{k-2}). \tag{10}$$

Since uS_za whence also uS_zb holds, we will be done if we show that $b \Vdash \mathsf{Z}_{k-1}$. To show that the remaining conjuncts of Z_{k-1} hold at b (that is $b \Vdash \Box C_{k-1} \wedge (E_{k-1} \rhd A_{k-2}) \wedge (E_{k-1} \rhd \mathsf{Z}_{k-2})$) simply observe that $\mathcal{G}_{2k-2}(v, a, b)$ and use (8), (9) and (10) to invoke the (IH) on v, a and b. $\Box$

Corollary 3.4. *If $F \models \mathcal{F}_{2k}$, then $F \models \widetilde{\mathsf{R}}_k$.*

Proof. Fix a forcing relation $\Vdash$ and let $w, x \in W$ such that $w \Vdash \mathsf{X}_k$ and $wRx \Vdash \mathsf{Y}_k$. Then for some y we have $xR^{C_k}y \Vdash A_k$. Thus there exists z with yS_wz and

$$z \Vdash B_k \wedge (\mathsf{X}_{k-1}) \tag{11}$$

(recall $\mathsf{X}_{-1} \equiv \top$). Since $F \models \mathcal{F}_{2k}$ we have $\mathcal{G}_{2k}(x, y, z)$. Thus by Lemma 3.3 we get

$$z \Vdash \Box C_k \wedge (E_k \rhd A_{k-1}) \wedge (E_k \rhd \mathsf{Z}_{k-1}) \ . \tag{12}$$

Combining (11) and (12) gives $z \Vdash \mathsf{Z}_k$. $\Box$

The reversal of this corollary is again preceded by a technical lemma. We shall denote by $\boldsymbol{a}_k$, $\boldsymbol{b}_k$, $\boldsymbol{c}_k$, and $\boldsymbol{e}_k$, propositional variables that shall play the role of the A_k, B_k, C_k and E_k respectively in the principles $\widetilde{\mathsf{R}}_n$. Likewise, by $\overline{\mathsf{X}}_k$ we shall denote the formula that arises by substituting $\boldsymbol{a}_j$ for A_j in X_k and $\boldsymbol{b}_j$ for B_j. The formulas $\overline{\mathsf{Y}}_k$ and $\overline{\mathsf{Z}}_k$ are defined similarly.

Lemma 3.5. *For any $k \geq 0$ and all $x, y, z \in W$. If for all forcing relations $\Vdash$ for which*

$$x \Vdash \overline{\mathsf{Y}}_k \text{ and } xR^{\boldsymbol{c}_k}_{\Vdash}y \text{ and } z \Vdash \overline{\mathsf{X}}_{k-1}$$

we also have

$$z \Vdash \Box \boldsymbol{c}_k \wedge (\boldsymbol{e}_k \rhd \boldsymbol{a}_{k-1}) \wedge (\boldsymbol{e}_k \rhd \overline{\mathsf{Z}}_{k-1}),$$

then $\mathcal{G}_{2k}(x, y, z)$.

Proof. Induction on k. Let $x, y, z \in W$ and assume the conditions of the lemma. Unfolding the definition of $\mathcal{G}_{2k}(x, y, z)$ shows us that we have to show that

1. for all u with zRu we have yS_xu $(k \geq 0)$;
2. and for all v and a with uS_xv and vRa we have uS_za $(k > 0)$;
3. and for all b with aS_zb we have $\mathcal{G}_{2(k-1)}(v, a, b)$ $(k > 0)$.

We will show 1 and 2 'by hand' and invoke the (IH) for 3. In each of the three cases we will choose similar but different forcing relations $\Vdash$.

We first show 1. So let zRu. Define

$$w \Vdash \boldsymbol{c}_k \Leftrightarrow yS_xw \quad \text{and} \quad w \Vdash \boldsymbol{a}_k \Leftrightarrow w = y.$$

And let all the other variables be false everywhere. Then $xR^{\boldsymbol{c}_k}_{\Vdash}y$ and $x \Vdash \neg(\boldsymbol{a}_k \rhd \neg \boldsymbol{c}_k)$. Since none of the $\boldsymbol{e}_i$ nor $\boldsymbol{a}_j$ with $j \neq k$ holds anywhere in the model, we trivially have $x \Vdash \overline{\mathsf{Y}}_k$ and $z \Vdash \overline{\mathsf{X}}_{k-1}$ and thus according to the conditions of the lemma in particular $z \Vdash \Box \boldsymbol{c}_k$. By definition of $\Vdash$ we thus have yS_xu which proves 1. Note that for $k = 0$ we only have to look after 1 hence we have now dealt with the base case of our induction.

Now we continue to show 2 assuming $k > 0$. Choose any v and a with uS_xv and vRa. As above define

$$w \Vdash \boldsymbol{c}_k \Leftrightarrow yS_xw \quad \text{and} \quad w \Vdash \boldsymbol{a}_k \Leftrightarrow w = y.$$

We now also define

$$
\begin{array}{ll}
w \Vdash \boldsymbol{e}_k \Leftrightarrow w = u & \text{and,} \\
w \Vdash \boldsymbol{a}_{k-1} \Leftrightarrow w = a \Leftrightarrow w \Vdash \boldsymbol{b}_{k-1} & \text{and,} \\
w \Vdash \boldsymbol{c}_{k-1} \Leftrightarrow aS_z w. &
\end{array}
$$

Let all the other propositional variables be false everywhere. Now $v \Vdash \overline{\mathsf{Y}}_{k-1}$ and thus $x \Vdash \overline{\mathsf{Y}}_k$. It is not hard to see that we also have $z \Vdash \overline{\mathsf{X}}_{k-1}$ and thus according to the condition of the lemma we have in particular $z \Vdash \boldsymbol{e}_k \rhd \boldsymbol{a}_{k-1}$. Since $zRu \Vdash \boldsymbol{e}_k$ there must be an a' with $uS_z a' \Vdash \boldsymbol{a}_{k-1}$. Since a is the only world that forces $\boldsymbol{a}_{k-1}$ we must have $uS_z a$.

To finish and show 3 choose b such that $aS_z b$. We want to show that $\mathcal{G}_{2(k-1)}(v, a, b)$. Invoking the (IH) it is enough to show that for any forcing relation $\Vdash$ for which

$$
v \Vdash \overline{\mathsf{Y}}_{k-1}, \quad \text{and} \quad vR_{\Vdash}^{\boldsymbol{c}_{k-1}} a \quad \text{and} \quad b \Vdash \overline{\mathsf{X}}_{k-2}, \tag{13}
$$

we also have

$$
b \Vdash (\boldsymbol{e}_{k-1} \rhd \boldsymbol{a}_{k-2}) \wedge (\boldsymbol{e}_{k-1} \rhd \overline{\overline{\mathsf{Z}}}_{k-2}) \wedge \Box \boldsymbol{c}_{k-1}. \tag{14}
$$

Our strategy in proving this is as follows. We slightly tweak $\Vdash$ to obtain $\Vdash'$. This $\Vdash'$ is similar to $\Vdash$ in that (13) still holds and moreover

$$
b \Vdash A \Leftrightarrow b \Vdash' A \quad \text{for subformulas } A \text{ of } (\boldsymbol{e}_{k-1} \rhd \boldsymbol{a}_{k-2}) \wedge (\boldsymbol{e}_{k-1} \rhd \overline{\overline{\mathsf{Z}}}_{k-2}) \wedge \Box \boldsymbol{c}_{k-1}. \tag{15}
$$

However, it is (possibly) different in that we now know that $x \Vdash' \overline{\mathsf{Y}}_k$, and $xR_{\Vdash'}^{\boldsymbol{c}_k} y$ and, $z \Vdash' \overline{\mathsf{X}}_{k-1}$ so that we may apply the main assumption of the lemma to $\Vdash'$ concluding $z \Vdash' \Box \boldsymbol{c}_k \wedge (\boldsymbol{e}_k \rhd \boldsymbol{a}_{k-1}) \wedge (\boldsymbol{e}_k \rhd \overline{\overline{\mathsf{Z}}}_{k-1})$. The latter will help us conclude (14).

Thus we consider an arbitrary forcing relation $\Vdash$ that satisfies (13). We modify $\Vdash$ to obtain $\Vdash'$ such that it satisfies

$$
\begin{array}{lcl}
w \Vdash' \boldsymbol{a}_k & \Leftrightarrow & w = y; \\
w \Vdash' \boldsymbol{e}_k & \Leftrightarrow & w = u; \\
w \Vdash' \boldsymbol{c}_k & \Leftrightarrow & yS_x w; \\
w \Vdash' \boldsymbol{a}_{k-1} & \Leftrightarrow & w = a; \\
w \Vdash' \boldsymbol{b}_{k-1} & \Leftrightarrow & w = b.
\end{array}
$$

Apart from these modifications, $\Vdash'$ will coincide with $\Vdash$. It is a straightforward check to see that we have (13) for $\Vdash'$ and that moreover (15) holds. In addition, by the definition of $\Vdash'$ we now also have

$$
x \Vdash' \overline{\mathsf{Y}}_k \quad \text{and} \quad xR_{\Vdash'}^{\boldsymbol{c}_k} y \quad \text{and} \quad z \Vdash' \overline{\mathsf{X}}_{k-1}. \tag{16}
$$

Thus, we see that $\Vdash'$ satisfies the antecedent of the condition of the lemma. Consequently, we have $z \Vdash' \boldsymbol{e}_k \rhd \overline{\overline{\mathsf{Z}}}_{k-1}$. Since $zRu \Vdash' \boldsymbol{e}_k$, there must exist some b' with $uS_z b' \Vdash' \overline{\overline{\mathsf{Z}}}_{k-1}$. But now, since $\boldsymbol{b}_{k-1}$ is a conjunct of $\overline{\overline{\mathsf{Z}}}_{k-1}$ and b is the only world that $\Vdash'$-forces $\boldsymbol{b}_{k-1}$, we must have $b \Vdash' \overline{\overline{\mathsf{Z}}}_{k-1}$. In particular, we conclude $b \Vdash' (\boldsymbol{e}_{k-1} \rhd \boldsymbol{a}_{k-2}) \wedge (\boldsymbol{e}_{k-1} \rhd \overline{\overline{\mathsf{Z}}}_{k-2}) \wedge \Box \boldsymbol{c}_{k-1}$; by (15) the same holds for $\Vdash$ and we are done. $\Box$

Putting this all together gives us the frame correspondence for $\widetilde{\mathsf{R}}_k$.

Theorem 3.6. *For any Veltman frame F and any natural number $k \geq 0$ we have*

$$F \models \mathcal{F}_{2k} \iff F \models \widetilde{\mathsf{R}}_k \iff F \models \mathsf{R}_{2k}.$$

Proof. The second equivalence is a direct consequence of Lemma 3.2 so we focus on the first equivalence.

The $\Rightarrow$ direction is just Corollary 3.4. For the other direction, fix some k, assume that $F \models \widetilde{\mathsf{R}}_k$ and let $wRxRyS_wz$. We have to show that $\mathcal{G}_{2k}(x, y, z)$. Now consider any forcing relation $\Vdash$ that satisfies $xR^{\mathbf{c}_k}_{\Vdash}y$, and $x \Vdash \overline{\mathsf{Y}}_k$ and, $z \Vdash \overline{\mathsf{X}}_{k-1}$. By Lemma 3.5 it is enough to show that

$$z \Vdash \Box \boldsymbol{c}_k \wedge (\boldsymbol{e}_k \rhd \boldsymbol{a}_{k-1}) \wedge (\boldsymbol{e}_k \rhd \overline{\mathsf{Z}}_{k-1}). \tag{17}$$

Now consider a forcing relation $\Vdash'$ where $\Vdash'$ is like $\Vdash$ except that

$$v \Vdash' \boldsymbol{a}_k \Leftrightarrow v = y \quad \text{and} \quad v \Vdash' \boldsymbol{b}_k \Leftrightarrow v = z.$$

Notice that $xR^{\mathbf{c}_k}_{\Vdash'}y$ and thus also $x \Vdash' \overline{\mathsf{Y}}_k$. But now we have $w \Vdash' \overline{\mathsf{X}}_k$ as well and thus $w \Vdash' \overline{\mathsf{Y}}_k \rhd \overline{\mathsf{Z}}_k$. Thus there must be some z' with $xS_wz' \Vdash \overline{\mathsf{Z}}_k$. Since $\boldsymbol{b}_k$ is a conjunct of $\overline{\mathsf{Z}}_k$ and z is the only world where $\boldsymbol{b}_k$ is forced we must have $z \Vdash' \overline{\mathsf{Z}}_k$. Since $\Box \boldsymbol{c}_k \wedge (\boldsymbol{e}_k \rhd \boldsymbol{a}_{k-1}) \wedge (\boldsymbol{e}_k \rhd \overline{\mathsf{Z}}_{k-1})$ does not involve $\boldsymbol{a}_k$ nor $\boldsymbol{b}_k$ we have (17). $\Box$

3.3 Arithmetical soundness

Via a series of lemmata we shall prove Theorem 3.7 to the effect that the hierarchy $\{\mathsf{R}_i\}_{i \in \omega}$ is arithmetically sound in any reasonable arithmetical theory.

Theorem 3.7. *Each of the R_i is arithmetically sound in any theory extending S^1_2.*

It is sufficient to prove that each of the R_{2m} is arithmetically sound in any reasonable arithmetical theory whence we shall focus on the principles $\widetilde{\mathsf{R}}_i$. We shall first exhibit a soundness proof of $\widetilde{\mathsf{R}}_1$ and then indicate how this is generalized to the rest of the hierarchy. And before proving $\widetilde{\mathsf{R}}_1$ we need some auxiliary lemmas.

Lemma 3.8. *Let T be any theory extending S^1_2. We have that for any arithmetical sentences E_1, A_0, B_0 and C_0 that*

$$T \vdash E_1 \rhd \neg(A_0 \rhd \neg C_0) \to \exists^{\mathsf{Cut}} J \, \Box \big(E_1 \to \forall^{\mathsf{Cut}} K {\in} \dot{J} \, \Diamond^{\dot{J}} (A_0 \wedge \Box^{\dot{K}} C_0)\big).$$

Proof. Reason in T and assume $E_1 \rhd \neg(A_0 \rhd \neg C_0)$. Note that by Lemma 2.6 we have $E_1 \rhd \forall^{\mathsf{Cut}} K \Diamond (A_0 \wedge \Box^{\dot{K}} C_0)$. Consequently, by Pudlák's Lemma, Lemma 2.2, we get $\exists J \, \big(E_1 \wedge \exists^{\mathsf{Cut}} K {\in} \dot{J} \, \Box^{\dot{J}} \neg (A_0 \wedge \Box^{\dot{K}} C_0) \rhd \bot\big)$. But this is provably the same as $\exists^{\mathsf{Cut}} J \, \Box \big(E_1 \to \forall^{\mathsf{Cut}} K {\in} \dot{J} \, \Diamond^{\dot{J}} (A_0 \wedge \Box^{\dot{K}} C_0)\big)$ as was to be shown. $\Box$

Lemma 3.9. *Let T be any theory extending S^1_2. We have that for any arithmetical sentences E_1, A_0, B_0 and C_0 that*

$$T \vdash \exists^{\mathsf{Cut}} J\ \Box\big(E_1 \to \forall^{\mathsf{Cut}} K{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0 \wedge \Box^{\dot{K}} C_0)\big) \to E_1 \rhd A_0.$$

Proof. Reason in T. From the assumption we get in particular that $\exists^{\mathsf{Cut}} J\ \Box\big(E_1 \to \Diamond^{\dot{J}} A_0\big)$ so that $\exists^{\mathsf{Cut}} J\ E_1 \rhd \Diamond^{\dot{J}} A_0 \rhd A_0$. ☐

Lemma 3.10. *Let T be any theory extending S^1_2. We have that for any arithmetical sentences E_1, A_0, B_0 and C_0 that*

$$T \vdash (A_0 \rhd B_0) \wedge \exists^{\mathsf{Cut}} J\ \Box\big(E_1 \to \forall^{\mathsf{Cut}} K{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0 \wedge \Box^{\dot{K}} C_0)\big) \to E_1 \rhd B_0 \wedge \Box C_0.$$

Proof. Reasoning in T we get from $\exists^{\mathsf{Cut}} J\ \Box\big(E_1 \to \forall^{\mathsf{Cut}} K{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0 \wedge \Box^{\dot{K}} C_0)\big)$ that $\forall^{\mathsf{Cut}} K\ \big(E_1 \rhd A_0 \wedge \Box^{\dot{K}} C_0\big)$. We combine this with $A_0 \rhd B_0 \to \exists^{\mathsf{Cut}} J\ \big(A_0 \wedge \Box^{\dot{J}} C_0 \rhd B_0 \wedge \Box C_0\big)$ to conclude $E_1 \rhd B_0 \wedge \Box C_0$. ☐

With these technical lemmas we can prove soundness of $\widetilde{\mathsf{R}}_1$.

Lemma 3.11. *Let T be any theory extending S^1_2. We have that for any arithmetical sentences E_1, A_1, B_1, A_0, B_0 and C_0 that*

$$\begin{array}{lll} T & \vdash & A_1 \rhd B_1 \wedge (A_0 \rhd B_0) \to \\ & & \quad \neg(A_1 \rhd \neg C_1) \wedge (E_1 \rhd \neg(A_0 \rhd \neg C_0)) \rhd \\ & & \qquad\qquad B_1 \wedge (A_0 \rhd B_0) \wedge \Box C_1 \wedge (E_1 \rhd A_0) \wedge (E_1 \rhd B_0 \wedge \Box C_0). \end{array}$$

Proof. We reason in T. Using our new technical lemma and Lemma 2.6 we get

$$\begin{array}{ll} \neg(A_1 \rhd \neg C_1) \wedge (E_1 \rhd \neg(A_0 \rhd \neg C_0)) & \to \\ \forall^{\mathsf{Cut}} K \Diamond (A_1 \wedge \Box^{\dot{K}} C_1)\ \wedge\ \exists^{\mathsf{Cut}} J\ \Box\big(E_1 \to \forall^{\mathsf{Cut}} L{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0 \wedge \Box^{\dot{L}} C_0)\big) & \to \\ \forall^{\mathsf{Cut}} K \Diamond \Big(A_1 \wedge \Box^{\dot{K}} C_1\ \wedge\ \exists^{\mathsf{Cut}} J{\in}\dot{K}\ \Box^{\dot{K}}\big(E_1 \to \forall^{\mathsf{Cut}} L{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0 \wedge \Box^{\dot{L}} C_0)\big)\Big). & \end{array}$$

The last step is due to the principle of outside-big inside-small (Lemma **??**) and allows us to conclude

$$\begin{array}{l} \forall^{\mathsf{Cut}} K\ \Big(\ \neg(A_1 \rhd \neg C_1) \wedge (E_1 \rhd \neg(A_0 \rhd \neg C_0))\ \rhd \\ \qquad\qquad A_1 \wedge \Box^{\dot{K}} C_1\ \wedge\ \exists^{\mathsf{Cut}} J{\in}\dot{K}\ \Box^{\dot{K}}\big(E_1 \to \forall^{\mathsf{Cut}} L{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0 \wedge \Box^{\dot{L}} C_0)\big)\ \Big). \end{array}$$

This can be combined with the fact that

$$A_1 \rhd B_1 \wedge (A_0 \rhd B_0) \to \exists^{\mathsf{Cut}} K\ \big(A_1 \wedge \sigma^{\dot{K}} \rhd B_1 \wedge \sigma \wedge (A_0 \rhd B_0)\ \big)$$

for this particular K holds for any $\sigma \in \Sigma_1$ to conclude

$$\begin{array}{l} A_1 \rhd B_1 \wedge (A_0 \rhd B_0) \to \neg(A_1 \rhd \neg C_1) \wedge (E_1 \rhd \neg(A_0 \rhd \neg C_0))\ \rhd \\ \qquad B_1 \wedge (A_0 \rhd B_0) \wedge \Box C_1\ \wedge\ \exists^{\mathsf{Cut}} J\ \Box\big(E_1 \to \forall^{\mathsf{Cut}} L{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0 \wedge \Box^{\dot{L}} C_0)\big). \end{array}$$

(Note that $\Box^{\dot{K}} C_1 \wedge \exists^{\mathsf{Cut}}\, J{\in}\dot{K}\ \Box^{\dot{K}}\ \big(E_1 \to \forall^{\mathsf{Cut}}\, L{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0{\wedge}\Box^{\dot{L}} C_0)\big)$ is equivalent to an $\exists\Sigma_1^b$ sentence relativized to $\dot{K}$.) Our technical lemmas 3.9 and 3.10 tell us that

$$\begin{aligned}(A_0 \rhd B_0) \wedge \exists^{\mathsf{Cut}} J\ \Box\big(E_1 \to \forall^{\mathsf{Cut}}\, L{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0 \wedge \Box^{\dot{L}} C_0)\big) \quad &\to \\ (E_1 \rhd A_0) \wedge (E_1 \rhd B_0 \wedge \Box C_0)&\end{aligned}$$

and we are done. $\Box$

The soundness proofs for $\widetilde{\mathsf{R}}_k$ is essentially not much different. We shall indicate where the soundness proof for $\widetilde{\mathsf{R}}_1$ needs to be modified and begin with modifications of the technical lemmas.

However, first we must inductively define a series of important formulas. In our definition we work with more variables than actually needed. However, we have chosen to do so since our variables can be interpreted as numbers or as formulas and we wish to avoid expressions like $\forall^{\mathsf{Cut}} J\ \Box\, \exists J{\in}\dot{J}\ \phi$.

$$\begin{aligned}\mathcal{H}_1 &:= \exists^{\mathsf{Cut}} J_1\ \Box\big(E_1 \to \forall^{\mathsf{Cut}} K_1{\in}\dot{J}_1\ \Diamond^{\dot{J}_1}(A_0 \wedge \Box^{\dot{K}_1} C_0)\big);\\ \mathcal{H}_{k+1} &:= \exists^{\mathsf{Cut}} J_{k+1}\ \Box\big(E_{k+1} \to \forall^{\mathsf{Cut}} K_{k+1}{\in}\dot{J}_{k+1}\ \Diamond^{\dot{J}_{k+1}}(A_k \wedge \Box^{\dot{K}_{k+1}} C_k \wedge \mathcal{H}_k^{\dot{K}_{k+1}})\big).\end{aligned}$$

It is easy to see that for each $k > 0$ the formula $\mathcal{H}_k$ is an $\exists\Sigma_1^b$ formula. The next lemmas show us that $\mathcal{H}_{k+1}$ are $\exists\Sigma_1^b$ consequences of the Σ_3 statements $E_{k+1} \rhd \mathsf{Y}_k$ which contain all the essential information for proving soundness. First we prove a simple modification of Lemma 3.8.

Lemma 3.12. *Let T be any theory extending S^1_2. We have that for any arithmetical sentences E_1, A_0, B_0 and C_0 and for any $\exists\Sigma_1^b$ formula σ that*

$$T \vdash E_1 \rhd \neg(A_0 \rhd \neg C_0) \wedge \sigma \to \exists^{\mathsf{Cut}} J\ \Box\ \big(E_1 \to \forall^{\mathsf{Cut}}\, K{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0 \wedge \Box^{\dot{K}} C_0 \wedge \sigma^{\dot{K}})\big).$$

Proof. We repeat the proof of Lemma 3.8. Note that, by our reading conventions the antecedent $E_1 \rhd \neg(A_0 \rhd \neg C_0) \wedge \sigma$ should be read as $E_1 \rhd \big(\neg(A_0 \rhd \neg C_0) \wedge \sigma\big)$. We reason in T and see that

$$\begin{aligned}\neg(A_0 \rhd \neg C_0) \wedge \sigma \quad &\to \quad \forall^{\mathsf{Cut}} K \Diamond(A_0 \wedge \Box^{\dot{K}} C_0) \wedge \sigma\\ &\to \quad \forall^{\mathsf{Cut}} K \Diamond(A_0 \wedge \Box^{\dot{K}} C_0 \wedge \sigma^{\dot{K}}).\end{aligned}$$

As before, the latter implies $\exists^{\mathsf{Cut}} J\ \Box\ \big(E_1 \to \forall^{\mathsf{Cut}}\, K{\in}\dot{J}\ \Diamond^{\dot{J}}(A_0{\wedge}\Box^{\dot{K}} C_0{\wedge}\sigma^{\dot{K}})\big)$. $\Box$

With this lemma we see that the $\mathcal{H}_{k+1}$ are an $\exists\Sigma_1^b$ encoding of information present in $E_{k+1} \rhd \mathsf{Y}_k$:

Lemma 3.13. *Let T be a theory containing S^1_2 and let the formulas $E_i, A_i,$ and C_i be arbitrary. For any number k we have that*

$$T \vdash E_{k+1} \rhd \mathsf{Y}_k \ \to\ \mathcal{H}_{k+1}.$$

Proof. By an external induction on k. For $k = 0$ this is simply Lemma 3.8. For the inductive case we reason in T and see that $E_{k+2} \rhd \mathsf{Y}_{k+1} \equiv E_{k+2} \rhd \neg(A_{k+1} \rhd \neg C_{k+1}) \wedge (E_{k+1} \rhd \mathsf{Y}_k)$. By the inductive hypothesis we have that $E_{k+1} \rhd \mathsf{Y}_k \to \mathcal{H}_{k+1}$ so that $E_{k+2} \rhd \mathsf{Y}_{k+1} \to E_{k+2} \rhd \neg(A_{k+1} \rhd \neg C_{k+1}) \wedge \mathcal{H}_{k+1}$. Since $\mathcal{H}_{k+1}$ is equivalent to an $\exists\Sigma^b_1$ formula, by Lemma 3.12 we see that

$$E_{k+2} \rhd \neg(A_{k+1} \rhd \neg C_{k+1}) \wedge \mathcal{H}_{k+1} \to \mathcal{H}_{k+2}$$

as was to be shown. $\square$

Moreover, the $\mathcal{H}_{k+1}$ formulas contain all the information to get the induction going as shown by the following lemma.

Lemma 3.14. *Let T be a theory containing S^1_2 and let the formulas $E_i, A_i, B_i,$ and C_i be arbitrary. For any number k we have that*

$$T \vdash (\mathsf{X}_k) \ \wedge \ \mathcal{H}_{k+1} \ \ \to \ \ E_{k+1} \rhd \mathsf{Z}_k.$$

Proof. By induction on k where the case $k = 0$ is just lemma 3.10. For the inductive case, we reason in T and assume $(\mathsf{X}_{k+1}) \ \wedge \ \mathcal{H}_{k+2}$.

From the definition of $\mathcal{H}_{k+2}$ we get

$$\exists^{\mathsf{Cut}} J_{k+2} \ \Box\big(E_{k+2} \to \forall^{\mathsf{Cut}} K_{k+2} {\in} \dot{J}_{k+2} \ \Diamond^{\dot{J}_{k+2}}(A_{k+1} \wedge \Box^{\dot{K}_{k+2}} C_{k+1} \wedge \mathcal{H}^{\dot{K}_{k+2}}_{k+1})\big)$$

so that $\exists^{\mathsf{Cut}} J_{k+2} \, \forall^{\mathsf{Cut}} K_{k+2} \ \Box\big(E_{k+2} \to \Diamond^{\dot{J}_{k+2}}(A_{k+1} \wedge \Box^{\dot{K}_{k+2}} C_{k+1} \wedge \mathcal{H}^{\dot{K}_{k+2}}_{k+1})\big)$

whence

$$\forall^{\mathsf{Cut}} K_{k+2} \ \big(E_{k+2} \rhd A_{k+1} \wedge \Box^{\dot{K}_{k+2}} C_{k+1} \wedge \mathcal{H}^{\dot{K}_{k+2}}_{k+1}\big). \tag{18}$$

From X_{k+1} –which is by definition equal to $A_{k+1} \rhd B_{k+1} \wedge (\mathsf{X}_k)$– we find via Pudlák's lemma, Lemma 2.2, a specific cut $\overline{K}_{k+2}$ such that for any formula σ in Σ_1 we obtain $A_{k+1} \wedge \sigma^{\overline{K}_{k+2}} \rhd B_{k+1} \wedge (\mathsf{X}_k) \wedge \sigma$. We can plug in this cut $\overline{K}_{k+2}$ to (18) to obtain via transitivity of $\rhd$ that

$$E_{k+2} \rhd B_{k+1} \wedge (\mathsf{X}_k) \wedge \Box C_{k+1} \wedge \mathcal{H}_{k+1}.$$

We are almost done but $B_{k+1} \wedge (\mathsf{X}_k) \wedge \Box C_{k+1} \wedge \mathcal{H}_{k+1}$ is not quite equal to Z_{k+1} as was needed. The missing conjuncts are $E_{k+1} \rhd A_k$ and $E_{k+1} \rhd \mathsf{Z}_k$. The first is easily seen to follow from $\mathcal{H}_{k+1}$ and the second follows from the inductive hypothesis applied to $(\mathsf{X}_k) \wedge \mathcal{H}_{k+1}$. $\square$

We are now ready to prove Theorem 3.7 that the whole hierarchy is arithmetically sound.

Theorem 3.15. *Let T be a theory containing S^1_2 and let A_i, B_i, C_i and E_i be arbitrary arithmetical formulas. We have for each number k that*

$$T \vdash \widetilde{\mathsf{R}}_k \quad \textit{id est} \quad T \vdash \mathsf{X}_k \to \mathsf{Y}_k \rhd \mathsf{Z}_k.$$

Proof. By an external induction on k where the base case is the soundness of $\widetilde{\mathsf{R}}_0$ which has been proven in [4]. Thus, we reason in T assuming $A_{k+1} \rhd B_{k+1} \wedge (\mathsf{X}_k)$. We need to conclude that $\mathsf{Y}_{k+1} \rhd \mathsf{Z}_{k+1}$. But Y_{k+1} is nothing but $\neg(A_{k+1} \rhd \neg C_{k+1}) \wedge (E_{k+1} \rhd \mathsf{Y}_k)$. By Lemma 3.13 we know that $(E_{k+1} \rhd \mathsf{Y}_k) \to \mathcal{H}_{k+1}$. Using this and reasoning as before we obtain

$$\begin{array}{rcl} \neg(A_{k+1} \rhd \neg C_{k+1}) \wedge (E_{k+1} \rhd \mathsf{Y}_k) & \to & \forall^{\mathsf{Cut}} K \Diamond (A_{k+1} \wedge \Box^{\dot{K}} C_{k+1}) \wedge (E_{k+1} \rhd \mathsf{Y}_k) \\ & \to & \forall^{\mathsf{Cut}} K \Diamond (A_{k+1} \wedge \Box^{\dot{K}} C_{k+1}) \wedge \mathcal{H}_{k+1} \\ & \to & \forall^{\mathsf{Cut}} K \Diamond (A_{k+1} \wedge \Box^{\dot{K}} C_{k+1} \wedge \mathcal{H}^{\dot{K}}_{k+1}). \end{array}$$

Consequently,

$$\forall^{\mathsf{Cut}} K \ \big(\neg(A_{k+1} \rhd \neg C_{k+1}) \wedge (E_{k+1} \rhd \mathsf{Y}_k) \rhd A_{k+1} \wedge \Box^{\dot{K}} C_{k+1} \wedge \mathcal{H}^{\dot{K}}_{k+1}\big).$$

This can be combined with Pudlák's Lemma on $A_{k+1} \rhd B_{k+1} \wedge (\mathsf{X}_k)$ to obtain

$$\neg(A_{k+1} \rhd \neg C_{k+1}) \wedge (E_{k+1} \rhd \mathsf{Y}_k) \rhd B_{k+1} \wedge (\mathsf{X}_k) \wedge \Box C_{k+1} \wedge \mathcal{H}_{k+1}.$$

It is easy to see that $\mathcal{H}_{k+1}$ implies $E_{k+1} \rhd A_k$. Moreover, Lemma 3.14 tells us that $(\mathsf{X}_k) \wedge \mathcal{H}_{k+1} \to E_{k+1} \rhd \mathsf{Z}_k$ so that we may conclude

$$\neg(A_{k+1} \rhd \neg C_{k+1}) \wedge (E_{k+1} \rhd \mathsf{Y}_k) \rhd B_{k+1} \wedge (\mathsf{X}_k) \wedge \Box C_{k+1} \wedge (E_{k+1} \rhd A_k) \wedge (E_{k+1} \rhd \mathsf{Z}_k)$$

as was to be shown. □

4 A broad series of principles

In this section we present a different series of principles. We refer to this series as the broad series since the frame-conditions –see Figure 2– are typically represented over a broader area than the slim hierarchy as discussed above.

4.1 A broad series

In order to define the second series we first define a series of auxiliary formulas. For any $n \geq 1$ we define the schemata U_n as follows.

$$\begin{aligned} \mathsf{U}_1 &:= \Diamond \neg (D_1 \rhd \neg C), \\ \mathsf{U}_{n+2} &:= \Diamond ((D_{n+1} \rhd D_{n+2}) \wedge \mathsf{U}_{n+1}). \end{aligned}$$

Now, for $n \geq 0$ we define the schemata R^n as follows.

$$\begin{aligned} \mathsf{R}^0 &:= A \rhd B \to \neg(A \rhd \neg C) \rhd B \wedge \Box C, \\ \mathsf{R}^{n+1} &:= A \rhd B \to \mathsf{U}_{n+1} \wedge (D_{n+1} \rhd A) \rhd B \wedge \Box C. \end{aligned}$$

As an illustration we shall calculate the first four principles.

$$\begin{array}{lcl} \mathsf{R}^0 & := & A \rhd B \to \neg(A \rhd \neg C) \rhd B \wedge \Box C \\ \mathsf{R}^1 & := & A \rhd B \to \Diamond \neg (D_1 \rhd \neg C) \wedge (D_1 \rhd A) \rhd B \wedge \Box C \\ \mathsf{R}^2 & := & A \rhd B \to \Diamond \Big[(D_1 \rhd D_2) \wedge \Diamond \neg (D_1 \rhd \neg C) \Big] \wedge (D_2 \rhd A) \rhd B \wedge \Box C \\ \mathsf{R}^3 & := & A \rhd B \to \Diamond \Big((D_2 \rhd D_3) \wedge \Diamond \Big[(D_1 \rhd D_2) \wedge \Diamond \neg (D_1 \rhd \neg C) \Big] \Big) \wedge (D_3 \rhd A) \\ & & \hfill \rhd B \wedge \Box C \end{array}$$

While the series R_i did define a hierarchy in that $\mathsf{R}_{i+1} \vdash \mathsf{R}_i$, we shall see that no such relation holds for the series R^i.

4.2 Frame conditions

It is not hard to determine the frame condition for the first couple of principles in this series and in Figure 2 we have depicted the first three frame-conditions. In this section we shall prove that the correspondence proceeds as expected. Informally, the frame condition for R^n shall be the universal closure of

$$x_{n+1}Rx_n \dots Rx_0Ry_0S_{x_1}y_1 \dots S_{x_n}y_nS_{x_{n+1}}y_{n+1}Ru \to y_0S_{x_0}u. \tag{19}$$

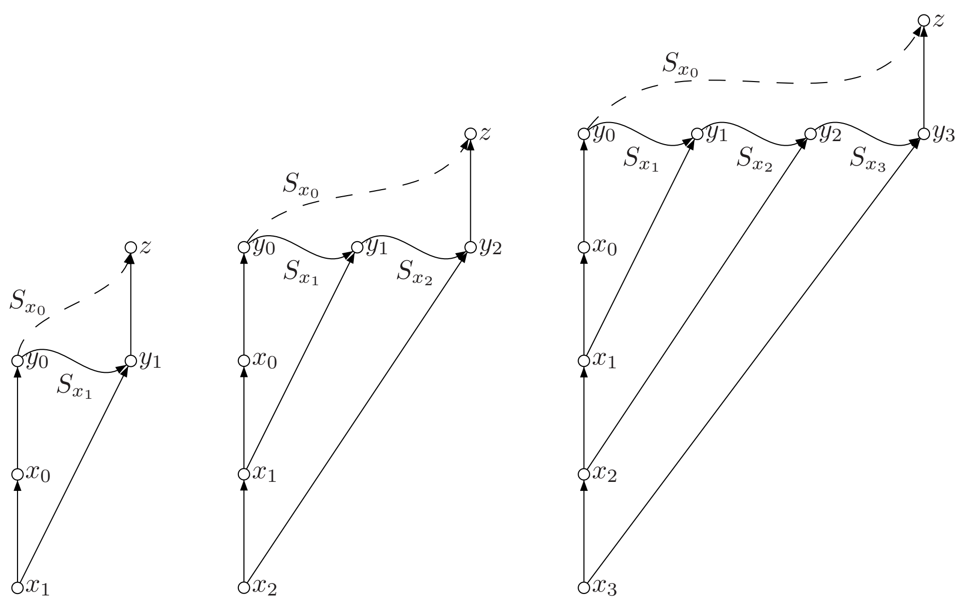

Figure 2: From left to right, this figure depicts the frame conditions $\mathcal{F}^0$ through $\mathcal{F}^2$ corresponding to R^0 through R^2. The reading convention is as always: if all the un-dashed relations are present as in the picture, then also the dashed relation should be there.

In order to make this frame condition precise and prove it, we shall first recast it in a recursive fashion. In writing (19) recursively we shall use those variables that will emphasize the relation with (19). Of course, free variables can be renamed at the readers liking.

First, we start by introducing a relation $\mathcal{B}_n$ that captures the antecedent of (19). Note that this antecedent says that first there is a chain of points x_i

related by R, followed by a chain of points y_i related by different S relations. The relation $\mathcal{B}_n$ will be applied to the end-points of both chains where the condition on the intermediate points is imposed by recursion.

$$\begin{aligned}\mathcal{B}_0(x_1, x_0, y_0, y_1) &:= x_1 R x_1 R y_0 S_{x_1} y_1,\\ \mathcal{B}_{n+1}(x_{n+2}, x_0, y_0, y_{n+2}) &:= \exists x_{n+1}, y_{n+1} \big(x_{n+2} R x_{n+1} \wedge \mathcal{B}_n(x_{n+1}, x_0, y_0, y_{n+1})\\ &\qquad\qquad\qquad\qquad \wedge y_{n+1} S_{x_{n+2}} y_{n+2}\big).\end{aligned}$$

For every $n \geq 0$ we can now define the first order frame condition $\mathcal{F}^n$ as follows.

$$\mathcal{F}^n := \forall x_{n+1}, x_0, y_0, y_{n+1} \ \big(\mathcal{B}_n(x_{n+1}, x_0, y_0, y_{n+1}) \Rightarrow \forall u\, (y_{n+1} R u \Rightarrow y_0 S_{x_0} u)\big).$$

Sometimes we shall write $x_{n+1}\mathcal{B}_n[x_0, y_0]\, y_{n+1}$ conceiving the quaternary relation $\mathcal{B}_n$ as a binary relation indexed by the pair x_0, y_0. In what follows we let $F = \langle W, R, S\rangle$ be an arbitrary Veltman-frame. The next lemma follows from an easy induction on n.

Lemma 4.1. *For each number n we have that $\mathcal{B}_n[x_0, y_0] \subseteq R$, that is, if $x_{n+1}\mathcal{B}_n[x_0, y_0]\, y_{n+1}$, then $x_{n+1} R y_{n+1}$.*

To prove that $F \models \mathcal{F}^n$ implies $F \models \mathsf{R}^n$ we first need a technical lemma.

Lemma 4.2. *Let $w \in W$ and $\Vdash$ be a forcing relation on F. If*

$$x_{k+1} \Vdash \mathsf{U}_{k+1} \wedge (D_{k+1} \rhd A),$$

then there exist x_0, y_0 and y_{k+1} such that $\mathcal{B}_k(x_{k+1}, x_0, y_0, y_{k+1})$, $x_0 R^C_{\Vdash} y_0$ and $y_{k+1} \Vdash A$.

Proof. Induction on k. If $k = 0$ then $\mathsf{U}_{k+1} = \Diamond\neg(D_1 \rhd \neg C)$ and the statement is easily checked. For the inductive case, we assume

$$x_{k+2} \Vdash \mathsf{U}_{k+2} \wedge (D_{k+2} \rhd A).$$

Recall that $\mathsf{U}_{k+2} := \Diamond((D_{k+1} \rhd D_{k+2}) \wedge \mathsf{U}_{k+1})$. Thus, there exists some x_{k+1} with $x_{k+2} R x_{k+1}$ and

$$x_{k+1} \Vdash (D_{k+1} \rhd D_{k+2}) \wedge \mathsf{U}_{k+1}.$$

Applying the (IH) (with D_{k+2} substituted for A) we find x_0, y_0 and y_{k+1} with $\mathcal{B}_k(x_{k+1}, x_0, y_0, y_{k+1})$, $x_0 R^C_{\Vdash} y_0$ and $y_{k+1} \Vdash D_{k+2}$. As $\mathcal{B}_k(x_{k+1}, x_0, y_0, y_{k+1})$ we get $x_{k+1} R y_{k+1}$ (Lemma 4.1). Since we had $x_{k+2} R x_{k+1}$ we see that $x_{k+2} R y_{k+1} \Vdash D_{k+2}$, and since $x_{k+2} \Vdash D_{k+2} \rhd A$, we find some y_{k+2} with $y_{k+1} S_{x_{k+2}} y_{k+2}$ and $y_{k+2} \Vdash A$. By definition of $\mathcal{B}_{k+1}$ we have $\mathcal{B}_{k+1}(x_{k+2}, x_0, y_0, y_{k+2})$. $\square$

Corollary 4.3. *If $F \models \mathcal{F}^n$ then $F \models \mathsf{R}^n$.*

Proof. Induction on n. For $n = 0$ this is known (see [4]), so we assume $n > 0$. Let $\Vdash$ be a forcing relation, let $x_{n+1}, x_n \in W$ and assume $x_{n+1} \Vdash A \rhd B$, $x_{n+1} R x_n$ and $x_n \Vdash \mathsf{U}_n \wedge (D_n \rhd A)$. By Lemma 4.2 we find x_0, y_0 and y_n such that $\mathcal{B}_{n-1}(x_n, x_0, y_0, y_n)$, with $x_0 R^C_{\Vdash} y_0$ and $y_n \Vdash A$. We have that $\mathcal{B}_{n-1}(x_n, x_0, y_0, y_n)$ implies $x_n R y_n$ (Lemma 4.1) and thus since $x_{n+1} R x_n$ we also we have $x_{n+1} R y_n \Vdash A$. By assumption $x_{n+1} \vdash A \rhd B$ so that for some y_{n+1} we have $y_n S_{x_{n+1}} y_{n+1} \Vdash B$. Clearly, we also have $x_n S_{x_{n+1}} y_{n+1}$ so that we are done if we have shown that $y_{n+1} \Vdash \Box C$. To this extent, we choose some u with $y_{n+1} R u$. Since we have that $\mathcal{B}_n(x_{n+1}, x_0, y_0, y_{n+1})$, by $\mathcal{F}^n$ we have also $y_0 S_{x_0} u$. But $x_0 R^C_{\Vdash} y_0$ and thus we have $u \Vdash C$, as required. $\Box$

To prove the converse implication, we start again with a technical lemma. As before we shall denote by $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}_k$, propositional variables that shall play the role of the A, B, C and D_k respectively in the principles R^n. Let $\overline{\mathsf{U}}_k$ denote the formula that arises by simultaneously substituting $\boldsymbol{c}$ for C and $\boldsymbol{d}_k$ for D_k in U_k.

Lemma 4.4. *Let $\{\boldsymbol{a}, \boldsymbol{c}, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_{k+1}\}$ be a collection of distinct propositional variables. If $F \models \mathcal{B}_k(x_{k+1}, x_0, y_0, y_{k+1})$, then there exists a forcing relation $\Vdash$ on F such that*

1. $x_{k+1} \Vdash \overline{\mathsf{U}}_{k+1} \wedge (\boldsymbol{d}_{k+1} \rhd \boldsymbol{a})$;
2. $x \Vdash \boldsymbol{c}$ *iff* $y_0 S_{x_0} x$;
3. $x \Vdash \boldsymbol{a} \Leftrightarrow x = y_{k+1}$;
4. $x \nVdash \boldsymbol{p}$ *for any* $\boldsymbol{p} \notin \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{k+1}, \boldsymbol{c}, \boldsymbol{a}\}$.

Proof. The idea is very simple using the informal description of $\mathcal{B}_k$ being the antecedent of (19). We define a valuation $\Vdash$ so that $\boldsymbol{d}_{i+1}$ is only true at y_i and $\boldsymbol{a}$ is only true at y_{k+1}. Moreover, we define $x \Vdash \boldsymbol{c}$ iff $y_0 S_{x_0} x$ and $x \nVdash \boldsymbol{p}$ for any $\boldsymbol{p} \notin \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{k+1}, \boldsymbol{c}, \boldsymbol{a}\}$. It is not hard to see that $x_{k+1} \Vdash \overline{\mathsf{U}}_{k+1} \wedge (\boldsymbol{d}_{k+1} \rhd \boldsymbol{a})$ for this valuation $\Vdash$.

To make the argument precise, we proceed by induction on k. If $k = 0$ then $\mathcal{B}_k(x_1, x_0, y_0, y_1)$ simply means $x_1 R x_0 R y_0 S_{x_1} y_1$ and we define

$$x \Vdash \boldsymbol{a} \Leftrightarrow x = y_1, \quad x \Vdash \boldsymbol{c} \Leftrightarrow y_0 S_{x_0} x \quad \text{and,} \quad x \Vdash \boldsymbol{d}_1 \Leftrightarrow x = y_0.$$

The lemma is easily checked if we further define $x \nVdash \boldsymbol{p}$ for any $\boldsymbol{p} \notin \{\boldsymbol{d}_1, \boldsymbol{c}, \boldsymbol{a}\}$.

For the inductive case we consider $k > 0$. Then $\mathcal{B}_k(x_{k+1}, x_0, y_0, y_{k+1})$ implies that there are x_k and y_k such that

$$x_{k+1} R\, x_k\, \mathcal{B}_{k-1}[x_0, y_0]\, y_k\, S_{x_{k+1}}\, y_{k+1}.$$

The (IH) (with $\boldsymbol{d}_{k+1}$ substituted for $\boldsymbol{a}$) gives a forcing relation $\Vdash$ such that

$$x_k \Vdash \overline{\mathsf{U}}_k \wedge (\boldsymbol{d}_k \rhd \boldsymbol{d}_{k+1}), \quad x_0 R^{\boldsymbol{c}}_{\Vdash} y_0, \quad x \Vdash \boldsymbol{d}_{k+1} \Leftrightarrow x = y_k$$

and $x \nVdash \boldsymbol{p}$ for $\boldsymbol{p} \notin \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{k+1}, \boldsymbol{c}\}$. So we have $x_{k+1} \Vdash \Diamond\big(\overline{\mathsf{U}}_k \wedge (\boldsymbol{d}_k \rhd \boldsymbol{d}_{k+1})\big)$; in other words $x_{k+1} \Vdash \overline{\mathsf{U}}_{k+1}$. We now define $\Vdash'$ as follows

$$x \Vdash' \boldsymbol{a} \Leftrightarrow x = y_{k+1} \quad \text{and} \quad x \Vdash' \boldsymbol{p} \Leftrightarrow x \Vdash \boldsymbol{p} \text{ for } \boldsymbol{p} \neq \boldsymbol{a}.$$

Clearly, the properties $x_k \Vdash \overline{\mathsf{U}}_k \wedge (\boldsymbol{d}_k \rhd \boldsymbol{d}_{k+1})$, $aR^{\boldsymbol{c}}_{\Vdash}b$, $x \Vdash \boldsymbol{d}_{k+1} \Leftrightarrow x = y_k$ simply extend to $\Vdash'$ and likewise we have that $x \nVdash' \boldsymbol{p}$ for any $\boldsymbol{p} \notin \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{k+1}, \boldsymbol{c}, \boldsymbol{a}\}$. Moreover, we now have $x_{k+1} \Vdash' \boldsymbol{d}_{k+1} \rhd \boldsymbol{a}$ as well. □

As a corollary to this lemma, we can now obtain the full the frame conditions for the principles R^n.

Theorem 4.5. *For each number n we have $F \models \mathcal{F}^n$ iff $F \models \mathsf{R}^n$.*

Proof. The $\Rightarrow$ direction is just Corollary 4.3 so we focus on the other direction. Thus, we suppose that $F \models \mathsf{R}^n$, consider any $x_{n+1}, x_0, y_0, y_{n+1} \in W$ with $\mathcal{B}_n(x_{n+1}, x_0, y_0, y_{n+1})$ and set out to show that for any u with $y_{n+1}Ru$ we have $y_0 S_{x_0} u$. We now apply Lemma 4.4 and simultaneously substitute $\boldsymbol{a}$ for $\boldsymbol{d}_{n+1}$ and $\boldsymbol{b}$ for $\boldsymbol{a}$ to see that there exists a forcing relation $\Vdash$ such that

$$x_{n+1} \Vdash \overline{\mathsf{U}}_{n+1}[\boldsymbol{d}_{n+1}/\boldsymbol{a}] \wedge (\boldsymbol{a} \rhd \boldsymbol{b}), \quad x \Vdash \boldsymbol{c} \Leftrightarrow y_0 S_{x_0} x \quad \text{and} \quad x \Vdash \boldsymbol{b} \Leftrightarrow x = y_{n+1}.$$

Since $n = 0$ is known, we assume $n > 0$. Thus, we find x_n with $x_{n+1}Rx_n$ and $x_n \Vdash \overline{\mathsf{U}}_{n-1} \wedge \boldsymbol{d}_n \rhd \boldsymbol{a}$ (note that $\overline{\mathsf{U}}_{n-1}[\boldsymbol{d}_{n+1}/\boldsymbol{a}] = \overline{\mathsf{U}}_{n-1}$). Using $F \models \overline{\mathsf{R}}^n$ we see that there must exist some x with $x \Vdash \boldsymbol{b} \wedge \Box\boldsymbol{c}$. But y_{n+1} is the only world that forces $\boldsymbol{b}$ thus necessarily $y_{n+1} \Vdash \Box\boldsymbol{c}$. By the choice of $\Vdash$ we thus have that if $y_{n+1}Ru$ then $y_0 S_{x_0} u$. □

Using the frame condition we readily see that the broad series of principles does not define a hierarchy.

Corollary 4.6. *For $n \neq m$ we have $\mathbf{IL}\mathsf{R}^n \nvdash \mathbf{IL}\mathsf{R}^m$.*

Proof. For each $m \neq n$ it is easy to exhibit a frame F so that $F \models \mathcal{F}^n$ but $F \not\models \mathcal{F}^m$. □

4.3 Arithmetical soundness

We will now see that all the principles R^n are arithmetically sound and begin with a simple lemma.

Lemma 4.7. *For any theory T extending S^1_2 and any natural number $n > 0$, we have that*

$$T \vdash \mathsf{U}_n \to \forall^{\mathsf{Cut}} K \, \Diamond(D_n \wedge \Box^{\dot{K}} C).$$

Proof. We proceed by induction on n and first consider $n = 1$. Thus, we reason in T and assume U_1, that is, $\Diamond\neg(D_1 \rhd \neg C)$. We conclude $\Diamond\forall^{\mathsf{Cut}} K \, \Diamond(D_1 \wedge \Box^{\dot{K}} C)$, whence $\forall^{\mathsf{Cut}} K \, \Diamond\Diamond(D_1 \wedge \Box^{\dot{K}} C)$ and also $\forall^{\mathsf{Cut}} K \, \Diamond(D_1 \wedge \Box^{\dot{K}} C)$ as was to be shown.

Next, we consider the inductive case, again reasoning in T and assuming U_{n+1} which is $\Diamond\big((D_n \rhd D_{n+1}) \wedge \mathsf{U}_n\big)$. By the (IH) we conclude from U_n that

$$\forall^{\mathsf{Cut}} J \, \Diamond \, (D_n \wedge \Box^{\dot{J}} C). \tag{20}$$

By Lemma 2.8 we obtain from $D_n \rhd D_{n+1}$ that

$$\forall^{\mathsf{Cut}} K \, \exists^{\mathsf{Cut}} J \ D_n \wedge \Box^{\dot{J}} C \rhd D_{n+1} \wedge \Box^{\dot{K}} C. \tag{21}$$

Combining $D_n \wedge \Box^{\dot{J}} C \rhd D_{n+1} \wedge \Box^{\dot{K}} C \to \big(\Diamond(D_n \wedge \Box^{\dot{J}} C) \to \Diamond(D_{n+1} \wedge \Box^{\dot{K}} C)\big)$ with (20) and (21) under a $\Diamond$ we conclude that

$$\begin{array}{rcl} \Diamond\big((D_n \rhd D_{n+1}) \wedge \mathsf{U}_n\big) & \to & \Diamond\big(\forall^{\mathsf{Cut}} K \, \Diamond \, (D_{n+1} \wedge \Box^{\dot{K}} C)\big) \\ & \to & \forall^{\mathsf{Cut}} K \, \Diamond\big(\Diamond \, (D_{n+1} \wedge \Box^{\dot{K}} C)\big) \\ & \to & \forall^{\mathsf{Cut}} K \, \Diamond \, (D_{n+1} \wedge \Box^{\dot{K}} C) \end{array}$$

as was to be shown. □

With this lemma, we can now prove the soundness of the series R^n.

Theorem 4.8. *For each natural number n we have that R^n is arithmetically sound in any theory T extending S^1_2.*

Proof. Since we already know that R^0 is sound, we consider $n > 0$. We reason in T, assume $A \rhd B$ and set out to prove $\mathsf{U}_n \wedge (D_n \rhd A) \rhd B \wedge \Box C$. By Pudlák's Lemma we get

$$\exists^{\mathsf{Cut}} J \ A \wedge \Box^{J} C \rhd B \wedge \Box C. \tag{22}$$

On the other hand, by the generalization of Pudlák's Lemma (Lemma 2.8) applied to $D_n \rhd A$ we obtain that $\forall^{\mathsf{Cut}} J \, \exists^{\mathsf{Cut}} K \ D_n \wedge \Box^{\dot{K}} C \rhd A \wedge \Box^{\dot{J}} C$ so that $\forall^{\mathsf{Cut}} J \, \exists^{\mathsf{Cut}} K \ \big(\Diamond(D_n \wedge \Box^{\dot{K}} C) \to \Diamond(A \wedge \Box^{\dot{J}} C)\big)$. By Lemma 4.7 we see that $\mathsf{U}_n \to \forall^{\mathsf{Cut}} K \Diamond(A \wedge \Box^{\dot{K}} C)$. Combining these last two observations, we see that $\mathsf{U}_n \wedge (D_n \rhd A) \to \forall^{\mathsf{Cut}} J \, \Diamond \, (A \wedge \Box^{\dot{J}} C)$ so that $\forall^{\mathsf{Cut}} J \ \mathsf{U}_n \wedge (D_n \rhd A) \rhd A \wedge \Box^{\dot{J}} C$. Combining this with (22) yields $\mathsf{U}_n \wedge (D_n \rhd A) \rhd B \wedge \Box C$ as was to be shown. □

5 On the core interpretability logic IL(All)

Apart from the principles mentioned earlier in this paper the literature has considered various other principles too. Some of those are

W: $A \rhd B \to A \rhd B \wedge \Box\neg A$

W*: $A \rhd B \to B \wedge \Box C \rhd B \wedge \Box C \wedge \Box\neg A$

P_0: $A \rhd \Diamond B \to \Box(A \rhd B)$

R: $A \rhd B \to \neg(A \rhd \neg C) \rhd B \wedge \Box C$

In [11], **IL**(All) was conjectured to be **IL**W. In [13] this conjecture was falsified and strengthened to a new conjecture, namely that **IL**W*, which is a proper extension of **IL**W, is **IL**(All). In [8] it was proven that the logic **IL**W*P$_0$ is a proper extension of **IL**W*, and that **IL**W*P$_0$ is a subsystem of **IL**(All) (we write **IL**W*P$_0$ instead of **IL**$\{$W*, P$_0\}$). This falsified the conjecture from [13]. In [8] it is also conjectured that **IL**W*P$_0$ is not the same as **IL**(All).

In [7] it is conjectured that **IL**W*P$_0$ =**IL**(All) and this conjecture was refuted in [4] by proving that the logic **IL**RW is a subsystem of **IL**(All) and a proper extension of **IL**W*P$_0$.

It is easy to see that $A \rhd \Diamond B \to \Box(A \rhd \Diamond B) \ \in$ **IL**P $\cap$ **IL**M. In [14] it was shown however that $A \rhd \Diamond B \to \Box(A \rhd \Diamond B) \ \notin$ **IL**(All) thereby lowering the upper bound **IL**(All) $\subseteq$ **IL**P $\cap$ **IL**M. Since $A \rhd \Diamond B \to \Box(A \rhd \Diamond B)$ is reminiscent of the modally incomplete principle P$_0$, we remark here that the principle

$$A \rhd B \to \neg(A \rhd \Diamond C) \rhd B \wedge \Box\neg C$$

implies $A \rhd \Diamond B \to \Box(A \rhd \Diamond B)$ so that it cannot be in **IL**(All) either.

The current paper raises the previously known lower bound of **IL**(All). However, it seems unlikely that this will be the end of the story and the two series presented here seem amenable for interactions. Just by mere inspection of the frame conditions we observe that

$$\begin{aligned}\mathcal{F}_n &= \forall w, x, y, z\ (\mathcal{B}_0(w, x, y, z) \Rightarrow \mathcal{G}_n(x, y, z)),\\ \mathcal{F}^n &= \forall w, x, y, z\ (\mathcal{B}_n(w, x, y, z) \Rightarrow \mathcal{G}_0(x, y, z)).\end{aligned}$$

suggesting possible interactions. For example, a combination of R^1 and R$_1$ could yield

$$A \rhd B \to (C \rhd A) \wedge \Diamond\neg(C \rhd \neg D) \wedge (E \rhd \Diamond F) \rhd B \wedge \Box D \wedge (E \rhd F).$$

We note that the two series presented in this paper only spoke of S relations that were imposed by the frame conditions. This suggests that a new conjecture can be formulated.

Let $\mathfrak{F}$ be a class of **IL**-frames. By **IL**$[\mathfrak{F}]$ we shall denote the interpretability logic corresponding to this class. That is,

$$\mathbf{IL}[\mathfrak{F}] \ := \ \{A \mid \forall F \in \mathfrak{F}\, \forall^{\mathsf{valuation}} V\ \langle F, V\rangle \models A\}.$$

We now define the class of frames $\mathfrak{All}$ to be the set of frames where any S relation that is implied both by the **IL**M and the **IL**P frame condition is present. To make this more precise, let P denote the first-order frame condition of P and let M denote the first-order frame condition of M. Let $F(x, y, z)$ denote any sentence –first or higher order– in the language $\{R, \{S_x\}_{x\in W}\}$. We write **IL**P $\models F(x, y, z) \to yS_xz$ to denote that for any Veltman frame $\mathcal{F}$ for which $\mathcal{F} \models P$ we also have $\mathcal{F} \models F(x, y, z) \to yS_xz$. Likewise, we shall speak of

$\mathbf{IL}\mathsf{M} \models F(x,y,z) \to yS_xz$. With this notation, we define

$$\begin{aligned}\mathfrak{All} \quad := \quad \{\mathcal{F} \mid \Big(&\mathbf{IL}\mathsf{P} \models (F(x,y,z) \to yS_xz) \;\&\; \mathbf{IL}\mathsf{M} \models (F(x,y,z) \to yS_xz) \Rightarrow \\ &\mathcal{F} \models (F(x,y,z) \to yS_xz)\Big)\}.\end{aligned}$$

The second author poses the new conjecture

Conjecture 5.1. $\mathbf{IL}(\mathrm{All}) = \mathbf{IL}[\mathfrak{All}]$.

It is easy to formulate the conjecture where the antecedent $F(x,y,z)$ is replaced by a set of sentences rather than a single sentence yet it seems hard to imagine that this is needed. Note that the conjecture only speaks of principles related to imposed S relations. For example, this will leave out a principle like $A \rhd B \to (\Diamond A \wedge \Box\Box C \rhd B \wedge \Box C)$ as formulated in [7].

References

[1] A. Berarducci. The interpretability logic of peano arithmetic. *Journal of Symbolic Logic*, 55(3):1059–1089, 1990.

[2] S. Buss. *Bounded arithmetic.* Bibliopolis, Napoli, 1986.

[3] D.H.J. de Jongh and F. Veltman. Provability logics for relative interpretability. In P.P. Petkov, editor, *Mathematical Logic, Proceedings of the Heyting 1988 summer school in Varna, Bulgaria*, pages 31–42. Plenum Press, Boston, New York, 1990.

[4] E. Goris and J. J. Joosten. A new principle in the interpretability logic of all reasonable arithmetical theories. *Logic Journal of the IGPL*, 19(1):14–17, 2011.

[5] P. Hájek and P. Pudlák. *Metamathematics of First Order Arithmetic.* Springer-Verlag, Berlin, Heidelberg, New York, 1993.

[6] J. J. Joosten. On formalizations of the Orey-Hájek characterization for interpretability. *Under preparation*, 2015.

[7] J. J. Joosten and A. Visser. The interpretability logic of all reasonable arithmetical theories. The new conjecture. *Erkenntnis*, 53(1-2):3–26, 2000.

[8] J.J. Joosten. Towards the interpretability logic of all reasonable arithmetical theories. Master's thesis, University of Amsterdam, 1998.

[9] V. Y. Shavrukov. The logic of relative interpretability over Peano arithmetic. Preprint, Steklov Mathematical Institute, Moscow, 1988. In Russian.

[10] A. Tarski, A. Mostowski, and R. Robinson. *Undecidable Theories.* North-Holland, Amsterdam, 1953.

[11] A. Visser. Preliminary notes on interpretability logic. Technical Report LGPS 29, Department of Philosophy, Utrecht University, 1988.

[12] A. Visser. Interpretability logic. In P.P. Petkov, editor, *Mathematical Logic, Proceedings of the Heyting 1988 summer school in Varna, Bulgaria*, pages 175–209. Plenum Press, Boston, New York, 1990.

[13] A. Visser. The formalization of interpretability. *Studia Logica*, 50(1):81–106, 1991.

[14] A. Visser. An overview of interpretability logic. In M. Kracht, M. de Rijke, and H. Wansing, editors, *Advances in modal logic '96*, pages 307–359. CSLI Publications, Stanford, CA, 1997.